

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-229505

(43)Date of publication of application : 25.08.1998

(51)Int.Cl. H04N 5/14
G06T 5/20
H04N 7/24

(21)Application number : 09-348873 (71)Applicant : LUCENT TECHNOL INC

(22)Date of filing : 18.12.1997 (72)Inventor : OKADA HIROYUKI

(30)Priority

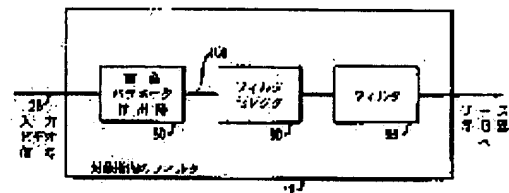
Priority number : 96 827911 Priority date : 18.12.1996 Priority country : US

(54) FILTERING DEVICE AND METHOD FOR LOW BIT RATE VIDEO SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide an object oriented filter to filter a video image before coding the video signal.

SOLUTION: An object oriented filter 12 is constituted by including an image parameter extract device 50, a filter selector 90 and a filter 95. The image parameter extract device 50 analyzes an input video sequence to conduct mapping of picture elements in relation to one or plural image parameters. After mapping, the filter selector 90 ranks the identified image parameter based on the importance to the entire video sequence and selects a filtering coefficient to adjust the strength of the filter used for filtering the picture element relating to the identified image parameters. Then the picture element relating to each image parameter is filtered by using the filter strength selected with the filter selector 90.



LEGAL STATUS

[Date of request for examination] 23.02.2000
[Date of sending the examiner's decision of rejection] 25.11.2002
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection] 2003-02911
[Date of requesting appeal against examiner's decision of rejection] 24.02.2003
[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-229505

(43) 公開日 平成10年(1998) 8月25日

(51) Int. Cl. ⁶	識別記号	P I	
H 0 4 N 5/14		H 0 4 N 5/14	Z
G 0 6 T 5/20		G 0 6 F 15/68	4 0 0 A
H 0 4 N 7/24		H 0 4 N 7/13	Z

審査請求 未請求 請求項の数22 O L (全 18 頁)

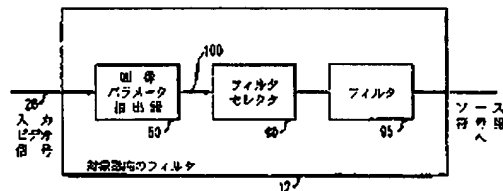
(21) 出願番号	特願平9-348373	(71) 出願人	596092698 ルーセント テクノロジーズ インコーポ レーテッド アメリカ合衆国, 07974-0636 ニュージ ャーシー, マレイ ヒル, マウンテン ア ヴェニュー 600
(22) 出願日	平成9年(1997)12月18日	(72) 発明者	オカダ ヒロユキ 千葉県千葉市緑区菅田町2-24-7-C 208
(31) 優先権主張番号	08/827911	(74) 代理人	弁理士 岡部 正夫 (外11名)
(32) 優先日	1996年12月18日		
(33) 優先権主張国	米国 (US)		

(54) 【発明の名称】 低ビットレートビデオシステムのためのフィルタリング装置および方法

(57) 【要約】 (修正有)

【課題】 ビデオの符号化の前にビデオ画像をフィルタリングするための対象指向のフィルタを提供する。

【解決手段】 この対象指向のフィルタ12は、画像パラメータ抽出器50、フィルタセクタ90及びフィルタ95を含んで構成される。画像パラメータ抽出器は、入力ビデオシーケンスを分析することで、一つあるいは複数の画像パラメータと関連する画素のマッピングを行ない、それを終えたら、フィルタセクタは、識別された画像パラメータを、ビデオシーケンスの全体に対する重要度に基づいてランク付けし、次に、これら識別された画像パラメータと関連する画素をフィルタリングするために使用されるフィルタの強度を調節するためのフィルタリング係数を選択する。その後、各画像パラメータと関連する画素が、フィルタセクタによって選択されたフィルタ強度を用いてフィルタリングされる。



(2)

特開平10-229505

1

【特許請求の範囲】

【請求項1】 ビデオ信号をフィルタリングするための装置であって、この装置が、ビデオ信号を受信するための第一のコネクタ；およびビデオ信号を分析すること
で、前記ビデオ信号内の少なくとも二つの所定の画像パラメータと関連する画素を識別するための抽出器を含み、ここで前記画像パラメータの一つと関連する画素の数が、前記ビデオ信号内の画素の総数より少なく、この装置がさらに前記所定の画像パラメータの少なくとも一つと関連する画素に第一の所定の係数を割当ててセレクト
10 タ；および前記ビデオ信号をフィルタリングするフィルタを含み、ここで前記第一の所定の係数を割当てられた画素に適用されるフィルタ強度と、前記ビデオ信号内の少なくとも一つの他の画素に適用されるフィルタ強度とが異なり、この装置がさらにフィルタリングされたビデオ信号を送信するための第二のコネクタを含むことを特徴とする装置。

【請求項2】 前記セレクトが、前記所定の画像パラメータを、前記ビデオシーケンス全体に対するそれらの重要度に基づいて、ランク付けすることを特徴とする請求項1の装置。

【請求項3】 前記セレクトが、前記第一の所定の係数に、 $1 < N < 256$ なる範囲内の値を割当ててことを特徴とする請求項1の装置。

【請求項4】 前記抽出器が、エッジパラメータと関連する画素を識別することを特徴とする請求項1の装置。

【請求項5】 前記エッジパラメータと関連する画素が、隣接画素に対する画素強度の大きさの変化に基づいて識別されることを特徴とする請求項4の装置。

【請求項6】 前記抽出器が、皮膚領域パラメータと関連する画素を識別することを特徴とする請求項1の装置。

【請求項7】 前記皮膚領域パラメータと関連する画素が、前記ビデオ信号内の対象の形状を分析することによって所望の形状の対象を識別することと、前記所望の形状の対象からの少なくとも一つの画素が皮膚領域であることを示す線度を持つか否かを決定することによって決定されることを特徴とする請求項6の装置。

【請求項8】 前記抽出器が、目鼻口（ENM）領域パラメータと関連する画素を識別することを特徴とする請求項1の装置。

【請求項9】 前記ENM領域パラメータと関連する画素が、顔を含む可能性が高い対象を識別するために、前記ビデオ信号内の対象の形状を分析することと、前記顔を含む可能性が高い識別された対象が対称の軸を持つか否かを決定することによって決定されることを特徴とする請求項8の装置。

【請求項10】 前記抽出器が、移動領域パラメータと関連する画素を識別することを特徴とする請求項1の装置。

2

【請求項11】 前記移動領域パラメータと関連する画素が、ブロックベースの移動推定技法を用いて識別されることを特徴とする請求項10の装置。

【請求項12】 ビデオ信号をフィルタリングする方法であって、この方法が、ビデオ信号を受信するステップ；および前記ビデオ信号内の少なくとも二つの所定の画像パラメータと関連する画素を識別するためのビデオ信号を分析するステップを含み、ここで前記画像パラメータの一つと関連する画素の数が、前記ビデオ信号内の画素の総数より少なく、この方法がさらに前記所定の画像パラメータの少なくとも一つと関連する画素に第一の所定の係数を割当ててステップ；および前記ビデオ信号をフィルタリングするステップを含み、ここで前記第一の所定の係数を割当てられた画素に適用されるフィルタ強度と、前記ビデオ信号内の少なくとも一つの他の画素に適用されるフィルタ強度とが異なり、この方法がさらに前記フィルタリングされたビデオ信号を送信するステップを含むことを特徴とする方法。

【請求項13】 前記所定の画像パラメータが、前記ビデオシーケンス全体に対するそれらの重要度に基づいてランク付けされることを特徴とする請求項12の方法。

【請求項14】 前記所定の画像パラメータの少なくとも一つの関連する画素に割当てられる前記第一の所定の係数が、 $1 < N < 256$ なる範囲内であることを特徴とする請求項12の方法。

【請求項15】 前記ビデオ信号内に識別される少なくとも二つの所定の画像パラメータの一つが、エッジパラメータであることを特徴とする請求項12の方法。

【請求項16】 前記エッジパラメータと関連する画素が、隣接画素に対する画素強度の大きさの変化に基づいて識別されることを特徴とする請求項15の方法。

【請求項17】 前記ビデオ信号内に識別される少なくとも二つの所定の画像パラメータの一つが、皮膚領域パラメータであることを特徴とする請求項12の方法。

【請求項18】 前記皮膚領域パラメータと関連する画素が、前記ビデオ信号内の対象の形状を分析することによって所望の形状の対象を検出することと、前記所望の形状の対象からの少なくとも一つの画素が皮膚領域であることを示す線度を持つか否かを決定することによって識別されることを特徴とする請求項17の方法。

【請求項19】 前記ビデオ信号内に識別される少なくとも二つの所定の画像パラメータの一つが、目鼻口（ENM）領域パラメータであることを特徴とする請求項12の方法。

【請求項20】 前記目鼻口（ENM）領域パラメータと関連する画素が、顔を含む可能性が高い対象を識別するために前記ビデオ信号内の対象の形状を分析することと、前記顔を含む可能性が高い識別された対象が対称の軸を持つか否かを決定することによって識別されることを特徴とする請求項19の方法。

(3)

特開平10-229505

3

【請求項21】 前記ビデオ信号内に識別される少なくとも二つの所定の画像パラメータの一つが、移動領域パラメータであることを特徴とする請求項12の方法。

【請求項22】 前記移動領域パラメータに関連する画像が、ブロックベースの移動推定技法を用いて識別されることを特徴とする請求項21の方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、マルチメディア用途のための低ビットレート通信システム、例えば、ビデオ電話会議システムに関する。より詳細には、ビデオ画像をビデオ符号化の前にフィルタリングする方法およびシステムに関する。

【0002】

【従来の技術】フルカラー、フルモーション画像のストレージおよび伝送の需要はますます増加している。これらの画像は、娯楽、例えば、動画あるいはテレビジョン製作のために使用されるのみでなく、分析および診断タスク、例えば、工学的分析および医療診断のために使用されている。これらの画像をデジタル形式にて提供することには幾つかの長所がある。例えば、デジタル画像は、画質を上げ、画像の操作がし易い。また、デジタルビデオ画像は、信号の劣化が小さく、何世代も正確に複製が可能である。

【0003】他方において、デジタルビデオは、ストレージのために多量のメモリ容量を必要とし、同等に伝送のためにも高帯域幅のチャネルを必要とする。例えば、256のグレーレベルを持つ512×512画素の単一のグレースケール画像で、256、000バイト以上のストレージが必要とされる。フルカラー画像になると、ほぼ800、000バイトが必要となる。自然に見える動画では、画像を一秒間に少なくとも30回更新することが必要とされる。自然に見えるフルカラー動画のための伝送チャネルでは、このために、約1.9億ビット/秒を収容できることが必要とされる。ただし、ビデオ電話、ビデオオンデマンド用セットトップボックス、ビデオ電話会議システムなどを含む現代のデジタル通信用途における伝送チャネルは、帯域幅に制約があり、ビデオ画像情報を伝送するために利用可能なビット数は、1.9億ビット/秒以下である。

【0004】このために、離散コサイン変換(DCT)などの幾つかの画像圧縮技術が、デジタルビデオ信号のストレージおよび伝送に必要とされる情報容量(バイト)を低減するために使用されている。これら技法は、一般的には、任意の自然の画像において見られる画像情報に含まれる多量の冗長に着目することで、デジタルビデオ画像を伝送、記録および複製するために使用される情報の量を低減する。例えば、ビデオシーケンス内のある対象(物体)が、同一の多くの画素を持つ場合、離散コサイン変換(DCT)技法においては、これら冗長画

4

素を識別するために0のデータ成分を使用することで、これら画素に関連する画像情報(バイト)を圧縮および伝送することが避けられる。離散コサイン変換技法においては、反対に、同一でない画素を識別するために、0でないデータ成分が使用される。こうして、伝送されるべきビデオ画像が、暗れた日の空の画像である場合は、離散コサイン変換(DCT)を使用する画像データ情報は、このような画像においては、現われる対象には変化が殆どまたは全く無いために、多数の0のデータ成分を含む。この結果、暗れた日の空の画像情報は、同一ではない画素を識別するために使用される0でない少数のデータ成分のみを伝送することで圧縮される。

【0005】画像圧縮技術、例えば、離散コサイン変換(DCT)に関連する一つの問題は、低ビットレート用途においては、DCT技術は、エラーによって乱された復号画像を生成する傾向があることである。一つのタイプの頻繁に発生するエラーは、“モスキートノイズ”と呼ばれるものである。これは、復号されたビデオセグメントにおいて、対象の近くに、これを取り巻くように“モスキート(蚊)”のような幻影を与えるように見える。“モスキートノイズ”は、典型的には、ビデオシーケンス内の対象のエッジおよび輪郭の所に出現する。“モスキートノイズ”などのノイズは、伝送するために利用可能なビットより、ビデオシーケンス内に符号化されるべき画像情報が多すぎる場合に発生する。例えば、ビデオシーケンス内に表現される対象のエッジおよび輪郭に対するDCTデータ情報は、これら領域は画像内に少しの冗長しか持たないために、多数の0でないデータ成分を含む。対象のエッジおよび輪郭領域に対する0でないデータ成分の数が、伝送のために利用可能なビット数より大きな場合、例えば、低ビットレートの電話システムの場合、対象に対するデータ成分の幾つかが正確に符号化できなくなる。このために、不正確に符号化されたデータ成分の幾つかが、伝送されたデータ成分が復号されたとき、“モスキートノイズ”を含むような画像に翻訳されることとなる。

【0006】典型的には、“モスキートノイズ”などのビデオ画像エラーは、ビデオ信号を画像圧縮の前にフィルタリングするプレフィルタの使用によって低減される。プレフィルタは、定義された周波数のレンジによって特徴づけられる。プレフィルタにビデオ信号が入力された場合、ビデオ信号に含まれる信号のうちの、そのプレフィルタに対して定義された周波数のレンジ内の周波数のみが、ビデオ符号器に出力される。ビデオ信号のうちの、そのプレフィルタに対して定義される周波数のレンジの外側の周波数は抑止される。こうして、プレフィルタリングは、本質的に、ビデオシーケンスからの画像情報の幾らかを除去し、これらシーケンスをより少ないビットを使用して符号化することができるようにし、結果として、“モスキートノイズ”などの符号化エラーを

(4)

特開平10-229505

5

6

低減する。

【0007】

【発明が解決しようとする課題】ただし、残念なことに、プレフィルタは、ビデオ信号を、単一の固定したレンジの周波数を使用してフィルタするために、画像情報が、ビデオシーケンスの内容にとって重要であると考えられる領域からも、重要でないと考えられる領域からも、一律に除去される。例えば、顔の領域を含むビデオシーケンスが、単一の固定されたレンジの周波数を使用してプレフィルタリングされた場合、顔の領域の、そのプレフィルタに対して定義された周波数の外側の周波数が抑止される。この結果、顔の領域は、しばしば、過度に平滑化された特徴（表情）を持つように描画され、元のビデオシーケンスには見られた顔に存在する細かな特徴、例えば、皺などがフィルタリングされる傾向があり、不自然な表情となる。

【0008】ビデオシーケンスの内の典型的に重要であると考えられる画像領域には、対象のエッジ、皮膚領域、顔領域、および移動する対象を取り巻く領域などが含まれる。反対に、ビデオシーケンスの内の重要ではないと考えられる画像領域には、画像の背景が含まれる。ビデオ信号のプレフィルタリングは、符号化に起因するエラーを低減することができるが、ただし、フィルタリング周波数の固定されたレンジを使用する方法では、フィルタ強度が単一となり、このために、ビデオシーケンスの幾つかの重要な領域からの画像情報が自動的に除去され、元の画像と同一の画像の復号されたバージョンとを比較した場合、画像品質が全般的に低下する。従って、符号化すべきビットの数を低減することが可能な画像圧縮技術を用いる改良されたプレフィルタリング構成が求められている。

【0009】

【課題を解決するための手段】本発明は、ビデオ画像をビデオの符号化の前にフィルタリングするための対象指向のフィルタに関する。一つの用途においては、この対象指向のフィルタは、ビデオ符号化／復号化（コーデック）装置のビデオ符号器との関連で使用される。この対象指向のフィルタは、最初に、入力ビデオ信号を分析することで、ビデオシーケンス内に含まれる一つあるいは複数の画像パラメータと関連する画素（ピクセル）の位置のマッピングを行なう。ここで使用される画像パラメータという用語は、ビデオシーケンスのある標相（アスペクト）と関連するパラメータを意味する。画像パラメータの例としては、対象のエッジ、皮膚領域、目鼻口（ENM）領域、および移動する対象を取り巻く領域などが含まれる。

【0010】いったん一つあるいは複数の画像パラメータと関連する画素のマッピングを終えると、この対象指向のフィルタは、ビデオシーケンスの各識別された画像パラメータに対するフィルタリング係数を選択する。こ

のフィルタリング係数は、その後各識別された画像パラメータと関連する、これら画素をフィルタリングするフィルタの強度を調節する。こうして、ビデオシーケンスの単一のフレーム内において、最初に、一つあるいは複数の画像パラメータと関連する画素がマッピングされ、その後、これらが、本発明の対象指向のフィルタによって、従来の技術のフィルタではビデオシーケンスの全ての画素が単一のフィルタ強度を使用してフィルタリングされたのとは対照的に、複数の異なるフィルタ強度を使用してフィルタリングされる。

【0011】これらのフィルタリング係数は、識別された画像パラメータを、それらパラメータのビデオシーケンス全体に対する重要度の順にランク付けすることによって選択される。例えば、あるビデオシーケンスが、野原を横切って歩く二人の人物を描画している場合、視聴者は彼あるいは彼女の注意を、背景（つまり、野原）ではなく、ビデオシーケンス内に含まれる特定の対象（つまり、歩いている二人の人物）に向ける傾向があるために、これら歩いている二人の人物の方が、ビデオシーケンスの内容にとって重要であるとみなされる。

【0012】本発明においては、ビデオシーケンスに含まれる一つあるいは複数の画像パラメータと関連する画素の位置がマッピングされ、その後これら画素をフィルタリングするフィルタの強度を調節するためのフィルタ係数が選択されるが、これは、これら画素が、ビデオシーケンス内のそれらの重要度に従ってフィルタリングされることとなるために有益である。

【0013】一つの実施例においては、対象指向のフィルタが、符号器、復号器、および符号化コントローラを含むビデオ符号化／復号化（コーデック）装置の他の要素と、独立して機能するが、統合される。一つの実施例においては、対象指向のフィルタは、入力ビデオ信号と符号器との間に挿入され、ビデオ画像の符号化の前に入力ビデオ信号をプレフィルタリングする。

【0014】本発明の一つの実施例においては、この対象指向のフィルタは、画像パラメータ抽出器、フィルタセクタ、およびフィルタを含む。画像パラメータ抽出器は、入力ビデオシーケンスを分析することで、少なくとも一つの画像パラメータと関連する画素の位置をマッピングする。画像パラメータ抽出器は、好ましくは、ビデオシーケンスの内容にとって重要であると思われるあらかじめ指定された画像パラメータを識別するようにプログラムされる。

【0015】いったん画像パラメータ抽出器が、選択された画像パラメータと関連する画素の位置をマッピングしたら、フィルタセクタによって、各画像パラメータのビデオシーケンス全体に対する相対的な重要度が決定される。フィルタセクタは、各画像パラメータの相対的な重要度を、所定の関係（アソシエーション）、例えば、階層に基づいて識別する。例えば、画像パラメータ

(5)

特開平10-229505

7

抽出器によって、目鼻口（ENM）領域、エッジ領域、および皮膚領域がマッピングされる場合、フィルタセクタは、ENM領域をエッジ領域よりも重要であるとしてランクし、エッジ領域を皮膚領域よりも重要であるとしてランクするようにプログラムされる。

【0016】識別された画像パラメータをそれらの相対的な重要度に従ってランク付けしたら、それら画像パラメータに関連する画素をフィルタリングするために用いられるフィルタの強度を調節するためのフィルタリング係数が、フィルタセクタによって選択される。フィルタセクタは、好ましくは、ビデオシーケンスの内容に対して重要であると思われる画像パラメータに関連する画素に最も弱いフィルタ強度が適用されるようにフィルタリング係数を選択する。ここで使用されるフィルタ強度という用語は、フィルタに対する周波数のレンジを意味する。弱いフィルタ強度は、広い範囲の周波数がフィルタを通過する。そして、このような弱いフィルタを使用してフィルタされた画素からは、非常に小さな画像情報（バイト）のみが除去される。反対に、強いフィルタ強度は、狭い範囲の周波数がフィルタを通過することを意味する。このために、強い強度を持つフィルタより弱い強度を持つフィルタの方が、より多くの画像情報（バイト）が通過する。

【0017】いったんフィルタ係数の選択を終えたら、識別された画像パラメータに関連する画素がフィルタによってフィルタリングされる。このフィルタは、好ましくは、分離可能なフィルタとされ、フィルタリング係数の関数の形式、例えば、

$$= \frac{1}{(2+N)^2} \begin{bmatrix} 1 & N & 1 \end{bmatrix} \times \begin{bmatrix} 1 & N & 1 \end{bmatrix}$$

のような形式を持つ。ここで、Nは、フィルタリング係数である。このようなフィルタでは、より強いフィルタ強度は、Nの小さな値と関連する。

【0018】本発明による対象指向のフィルタは、その後、ビデオシーケンスの後続のフレームを分析することで、画像パラメータを識別し、それらのビデオフレームをフィルタリングするのに適するフィルタリング係数を選択する。この対象指向のフィルタは、オプションとして、ビデオシーケンスの一つのフレームに対して選択されたフィルタリング係数を使用して、ビデオシーケンスの後続のフレームをフィルタすることも可能である。

【0019】本発明の他の目的および特徴が以下の詳細な説明を付録の図面を参照しながら読むことによって明らかになるものである。ただし、これらの図面は、単に解説を目的とするものであり、本発明の範囲を限定することを目的とするものではない。本発明の範囲については、特許請求の範囲を参照されるべきである。

【0020】

8

【発明の実施の形態】図1は、対象指向のフィルタ12が、ビデオ符号化／復号化システム、例えば、ビデオコーデック10（符号器／復号器）との関連で使用される本発明の一例としての用途を示す。ビデオ符号化／復号化システム、例えば、ビデオコーデック10は、主として、電話会議産業において、ビデオ画像シーケンスを、画像圧縮技術に基づいて、符号化および復号化するために用いられている。ビデオ画像シーケンスを符号化および復号化するために有効な画像圧縮技術の一例として、ITU-T勧告H.263（「狭帯域チャネルのためのビデオ符号化」）において説明されている離散コサイン変換（DCT）法を含む。本発明は、勿論、ビデオ符号器／復号器（コーデック）以外のビデオシステム、例えば、動画編集装置と共に使用することもできることに注意する。実際、本発明は、デジタルカラービデオ信号を入力として持つ任意の装置と共に使用することが可能である。

【0021】図1は、本発明の一つの実施例を示す。より詳細には図1は、ビデオコーデック10内に配置された本発明の対象指向のフィルタ12を点線で囲んで示す。対象指向のフィルタ12は、ビデオコーデック10の他の構成要素と統合されるが、ただし、独立して機能する。例えば、ビデオコーデック10は、追加の構成要素、例えば、符号化コントローラ16、ビデオ符号器22、およびビデオ復号器24を含む。これら構成要素については、後に、ビデオコーデック10の動作との関連で説明される。

【0022】対象指向のフィルタ12は、図2のブロック図により詳細に示すように、画像パラメータ抽出器50、フィルタセクタ90、およびフィルタ95を含む。これら画像パラメータ抽出器50、フィルタセクタ90、およびフィルタ95によって表される機能は、オプションとして、共有あるいは専用のハードウェアを使用して実現することが可能であり、またソフトウェアを実行する能力のあるハードウェアも含まれる。画像パラメータ抽出器50、フィルタセクタ90、およびフィルタ95の機能は、例えば、オプションとして、単一の共有プロセッサによって提供することも、あるいは複数の個別のプロセッサによって提供することも可能である。

【0023】さらに、画像パラメータ抽出器50、フィルタセクタ90、およびフィルタ95を表す個別の機能ブロックの使用は、ソフトウェアを実行することができるハードウェアを排他的に示すものではないことに注意する。上に説明される機能ブロックを実行するための追加の実施例においては、ハードウェアとしてのAT&T DSP16あるいはDSP32Cなどのデジタル信号プロセッサ（DSP）、後に説明される動作を遂行するためのソフトウェアを格納するための読出専用メモリ（ROM）、およびデジタル信号プロセッサ（DSP）

(6)

特開平10-229505

9

の結果を格納するためのランダムアクセスメモリ（RAM）が含まれる。オプションとして、ハードウェアとして大規模集積（VLSI）を用いる実現や、カスタムVLSI回路を汎用目的デジタル信号プロセッサ（DSP）回路と共に使用する方法なども考えられる。これらの任意の全ての実施例が、画像パラメータ抽出器50、フィルタセクタ90、およびフィルタ95として示される機能ブロックを実現するために使用できるものである。

【0024】本発明は、ビデオ画像シーケンスの符号化の前にビデオ画像をフィルタすることに関する。画像パラメータ抽出器50は、最初に、入力ビデオ画像を分析することで、ビデオシーケンス内に含まれる一つあるいは複数の画像パラメータと関連する画素をマッピングする。これら画像パラメータと関連する画素をマッピングすることは、重要であるとみなされる画像パラメータを決定するのに利点がある。例えば、ビデオ会議状況においては、視聴者は、彼あるいは彼女の注意を、画面の一人あるいは複数の人物の頭あるいは顔に向ける傾向がある。ビデオフレームの分析によって、皮膚領域が識別された場合、これは、高い確率で頭あるいは顔が存在することを示す。

【0025】その後、フィルタセクタ90によって、識別された画像パラメータと関連する画素は、各識別された画像パラメータのビデオシーケンスの内容全体についての相対的重要性に基づいて、ランクが付けられる。フィルタセクタは、識別された画像パラメータのランクに基づいて、後で各識別された画像パラメータと関連する画素をフィルタリングするフィルタの強度を調節するためのフィルタリング係数を決定する。結果として、各識別された画像パラメータに対して異なるフィルタリング係数が選択され、各画像パラメータと関連する画素は、異なる強度のフィルタを使用してフィルタリングされる。ビデオシーケンスの内容全体にとって重要であるとみなされる画像パラメータと関連する画素は、より弱い強度のフィルタを使用してフィルタリングされ、これら位置が過剰フィルタリングされるのが回避される。従来の技術によるフィルタでは、全ての画素が単一強度のフィルタを使用してフィルタされたが、これとは対照的に、本発明による方法は、選択された画像パラメータと関連する画素が異なる強度のフィルタを使用してフィルタリングされるために、対象指向のフィルタリングである。

【0026】以下では、図2を参照しながら、対象指向のフィルタ12の動作の説明の一部として、画像パラメータ抽出器50、フィルタセクタ90、およびフィルタ95の構成要素について説明する。対象の画像に対応するフレームのシーケンスを時間の関数として表す入力ビデオ信号26が、従来のビデオカメラ（図示せず）、例えば、シャープ社によって製造されるView Cam

10

から対象指向のフィルタ12に供給される。

【0027】図3に示される画像パラメータ抽出器50は、入力ビデオ信号26を分析することによって、ビデオフレーム内に含まれる、ビデオシーケンスの内容によって重要であると考えられる、一つあるいは複数の画像パラメータと関連する画素を識別する。いったん識別された一つあるいは複数の画像パラメータと関連する画素が決定されると、画像パラメータ位置信号100が生成され、フィルタセクタ90に供給される。オプションとして画像パラメータ抽出器50によって識別される画像パラメータの幾つかの例としては、対象のエッジ、皮膚領域、目鼻口（ENM）領域、移動する対象の周囲の領域、背景領域などが含まれる。

【0028】オプションとして識別される画像パラメータが対象のエッジである本発明の一つの実施例においては、画像パラメータ抽出器50は、図3に示すようにエッジ検出器51を含む。エッジ検出器51は、好ましくはビデオ画像の領域を分析することで、ビデオフレーム内に含まれる対象のエッジと関連する画素を識別する。エッジ検出器51は、図4に示すような前処理回路52と一体となっている。そして、前処理回路52は、一時ダウンサンブラ53、フィルタ54、デシメータ55、エッジ検出器56、および閾値回路57を含む。

【0029】一時ダウンサンブラ53は、エッジ検出のために用いられるビデオ信号のフレームの数を、分析のために入力ビデオ信号26から供給されるフレームの数を制限することで制限する機能をする。一例として、ビデオ信号、例えば、入力ビデオ信号26に対する典型的なフレームレートは、約30フレーム/秒（fps）であり、一連の各フレームは、本質的に前のフレームと同一の情報を含む。一連のフレームが本質的に同一の情報を含むために、エッジ検出のために、ビデオ信号からの少数のフレームのみを選択することによって計算の複雑さを低減することに利点がある。こうして、本発明においては、ダウンサンブラは、計算の複雑さを低減するために、エッジ検出のために、入力ビデオ信号のフレームを、4つに一つのみ選択する。この結果として、ダウンサンブラは、エッジ検出器に入力されるフレームレートを、約30フレーム/秒（fps）なるレートから、約7.5フレーム/秒（fps）なるレートに低減する。

【0030】フィルタ54は、典型的には、 352×288 画素なるサイズおよび π/c なる遮断周波数を持つ、ダウンサンプリングされたビデオフレームの空間フィルタリングを遂行するための、分解可能なフィルタとされる。ここで、 c は、好ましくは、後に説明されるデシメーション（分割）係数に等しくされる。典型的には、フィルタ、例えばフィルタ54は、周波数レンジを定義する。信号、例えば、ダウンサンプリングされた入力ビデオ信号26が、フィルタ54に供給されたとき、そのビデオ信号に含まれる周波数の内の、フィルタに対

(7)

待間平10-229505

11

して定義された周波数のレンジ内の周波数のみが出力される。反対に、ビデオ信号に含まれる周波数の内の、フィルタに対して定義されたレンジ外の周波数は、抑止される。フィルタ54の例としては、有限インパルス応答(FIR)フィルタ、および無限インパルス応答(IIR)フィルタが含まれる。

【0031】前処理回路52のフィルタ54は、対象指向のフィルタ12のフィルタ95と混同されるべきではない。フィルタ54の目的は、入力ビデオ信号26の一部分をフィルタリングすることにより、ビデオシーケンスに含まれる対象のエッジに関連する画素を検出することにより、入力ビデオ信号26のビデオ信号周波数を永久的に抑止することはない。これとは対照的に、対象指向のフィルタ12のフィルタ95は、入力ビデオ信号26をフィルタリングすることで、ビデオシーケンスの符号化の前に、選択された画像パラメータに関連する画素のビデオ信号周波数を永久的に抑止する。

【0032】こうしてフィルタされたビデオ信号は、デシメータ55に入り、ここで、ビデオ画像フレームの水平および垂直の両方の寸法が、エッジ分析のために、より小さな所定のサイズを持つ画像領域に分割される。一例として、デシメータ、例えばデシメータ55が、 $c=8$ なるデシメーション係数を待ち、ビデオ画像フレームが 352×288 画素なる寸法を持つ場合は、ビデオ画像フレームは 44×36 画素の寸法を持つ画像領域に分割される。

【0033】エッジ検出器56は、ビデオ画像フレームの分割された各画像領域に関してエッジ検出を遂行することで、対象のエッジを探す。任意のビデオ画像フレーム内の対象のエッジは、典型的には、隣接画素に対する画素強度の大きさの変化によって検出される。例えば、 3×3 なるサイズの画素の画像領域が、対象のエッジを含まない場合は、このような画像領域における隣接画素に対する画素強度の大きさは、ほぼ等しく、一例としてマトリックスAのようになる：

【数2】

*

$$\delta_{xA} = \begin{vmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{vmatrix} * \begin{vmatrix} 11 & 10 & 10 \\ 10 & 10 & 10 \\ 10 & 10 & 11 \end{vmatrix}$$

このたたみこみの結果は、ある程度、以下のような結果を与える：

12

$$A = \begin{vmatrix} 11 & 10 & 10 \\ 10 & 10 & 10 \\ 10 & 10 & 11 \end{vmatrix}$$

これとは対照的に、類似の 3×3 なるサイズの画像領域が、対象のエッジを含む場合は、このような画像領域における隣接画素に対する画素強度の大きさは、鋭い遷移を含み、一例としてマトリックスBのようになる：

【数3】

$$B = \begin{vmatrix} 10 & 50 & 90 \\ 50 & 50 & 90 \\ 90 & 90 & 90 \end{vmatrix}$$

【0034】エッジ検出器、例えばエッジ検出器56は、対象のエッジを、Sobelオペレータ技法を含む幾つかの技法を利用して、二次元Sobelオペレータ、例えば、 δ_x および δ_y のたたみこみの二乗を総和することによって、識別する。ここで、ある分割された画像領域における隣接画素に対する画素強度の大きさは、例えば、マトリックスAあるいはBに示されるような値を持つ。Sobelオペレータ技法を使用する一つの実施例においては、水平オペレータ δ_x および垂直オペレータ δ_y によって形成される二次元にて表現されるSobelオペレータ：

【数4】

$$\delta_x = \begin{vmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{vmatrix} \quad \delta_y = \begin{vmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{vmatrix}$$

が対象のエッジを含まない画像領域における隣接画素に対する画素強度の大きさ、例えば、マトリックスAに示されるような隣接画素に対する画素強度の大きさと、たたみこまれた場合は、以下のようになり：

【数5】

$$\delta_{xA} = \begin{vmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{vmatrix} * \begin{vmatrix} 11 & 10 & 10 \\ 10 & 10 & 10 \\ 10 & 10 & 11 \end{vmatrix}$$

【数6】

(8)

特開平10-229505

$$\begin{aligned}
 \delta_{13} &= (-1 \times 11) + (0 \times 10) + (1 \times 10) + (-2 \times 10) + (0 \times 10) + (2 \times 10) + (-1 \times 10) \\
 &\quad + (0 \times 10) + (1 \times 11) = 0 \\
 \delta_{14} &= (1 \times 11) + (2 \times 10) + (1 \times 10) + (0 \times 10) + (0 \times 10) + (0 \times 10) + (-1 \times 10) \\
 &\quad + (-2 \times 10) + (-1 \times 11) = 0
 \end{aligned}$$

【0035】この場合、これらの大きさは、2つの次元において、概ねりとなる。これとは対照的に、Sobelオペレータが、対象のエッジを含む画像領域における隣接画素に対する画素強度の大きさ、例えば、マトリックス* Bに示されるような隣接画素に対する画素強度の大きさ、とたたみこまれた場合は、このたたみこみの結果は、ある程度、以下のような結果を与える：

$$\begin{aligned}
 \delta_{15} &= (-1 \times 10) + (0 \times 50) + (1 \times 90) + (-2 \times 50) + (0 \times 50) + (2 \times 90) + (-1 \times 90) \\
 &\quad + (0 \times 90) + (1 \times 90) = 160 \\
 \delta_{16} &= (1 \times 10) + (2 \times 50) + (1 \times 90) + (0 \times 50) + (0 \times 50) + (0 \times 90) + (-1 \times 90) \\
 &\quad + (-2 \times 90) + (-1 \times 90) = -160
 \end{aligned}$$

この場合は、これらの大きさは、概ねりとはならない。例えば、上に説明のSobelオペレータ技法を利用する、エッジ検出技法が、ビデオフレームの分割された各44×36画素領域に対して実行される。

【0036】次に、閾値回路57によって、各44×36画素の分割領域について、それら画素の隣接画素に対する画素強度のたたみこみされ総和された絶対値が、指定された値より大きいとかが識別され、閾値より大きいとして識別された画素に対しては、非ゼロの数値が割当てられる。他方、閾値回路57によって指定される値より小さな、隣接画素に対する画素強度のたたみこみされ総和された絶対値を持つ画素に対しては、ゼロの数値が割当てられる。次に、非ゼロの画素値に対応するエッジデータ信号58が、閾値回路57によって生成され、フィルタセクタ90に供給される。

【0037】閾値回路、例えば閾値回路57を組み込むことで、長所として、エッジでないビデオシーケンスに含まれる対象の輪郭領域が、エッジ検出器51によって、誤ってエッジであると識別することが阻止される。これは、隣接画素に対する画素強度における変化が小さな場合は、典型的には、閾値回路の指定された値より小さな、たたみこみされたおよび総和された絶対値が生成されるためである。

【0038】オプションとして識別される画像パラメータが皮膚領域である本発明の一つの実施例においては、画像パラメータ抽出器50は、図3に示すように、皮膚検出器60を含む。皮膚検出器60は、皮膚領域と関連する画素を識別する。これは、最初に、ビデオシーケンス内の対象の形状を分析することにより、皮膚領域を含む可能性の高い一つあるいは複数の対象を検出することから成る。次に、これら皮膚領域を含む可能性が高い領域を分析され、それら任意の一つあるいは複数の対象の画像

成分（画素）が、皮膚領域に特徴的な信号エネルギーを持つと否か決定される。その後、皮膚領域に特徴的な信号エネルギーを持つ画素を持つ対象の色調がサンプリングされ、その対象に対する皮膚色調値のレンジが決定される。次に、分析された対象に対するサンプリングされた皮膚色調値のレンジが、ビデオ画像内に含まれる全ての色調と比較され、ビデオシーケンス内の同一の皮膚色調値を持つ他の領域が識別される。皮膚検出器60は、図5に示すように、形状検出器61および色調検出器62を含む。

【0039】形状検出器61は、最初に、ビデオフレーム内の一つあるいは複数の皮膚領域と見られる領域を、ビデオフレーム内の対象のエッジの識別および任意のこのエッジが所定の形状の輪郭に類似するかどうかに基づいて検出する。皮膚領域である可能性が高い形状の例としては、楕円、弧、湾曲などが含まれる。この明細書において使用される湾曲という用語は、少なくとも一部分が直線ではないエッジを持つ形状を意味する。

【0040】形状検出器61の構成要素には、図6に示すように、形状検出前置プロセッサ64、粗スキャナ65、微細スキャナ66、および形状適合器67が含まれる。形状適合器67は、形状検出信号68を生成し、この信号は、色調検出器62に供給される。

【0041】形状検出前置プロセッサ64は、図4に示す前置処理回路52と同様の構成要素を含み、入力ビデオ信号26の領域を分析することで、ビデオフレーム内の対象のエッジを識別する。形状検出前置プロセッサ64の機能は、前に説明された前置処理回路52と同一であり、前置処理回路52との関連で前に説明されたので、皮膚検出を遂行するために必要な全てのエッジ情報は、決定されるものとする。また、形状検出前置プロセッサ64を形状検出器61の構成要素として含むこと

15

は、画像パラメータ抽出器50が、エッジ検出器、例えばエッジ検出器51、および皮膚検出器、例えば皮膚検出器60の両方を含む場合は、エッジ情報は、エッジ検出器51の前置処理回路52から得ることができるために、オプションである。

【0042】図6の説明に戻り、形状検出前置プロセッサ64によって生成された（あるいは、オプションとしてエッジ検出器51の前置処理回路52から得られた）エッジデータ信号は、形状検出器61の粗スキャナ65 10に入力される。粗スキャナ65は、エッジデータ信号を、 $B \times B$ 画素なるサイズ、例えば 5×5 画素なるサイズのブロックに分割する。各ブロックが、次に、粗スキャナ65によって、上に説明のエッジ分析に基づいて、そのブロック内の画素の少なくとも一つが、非ゼロの値を持つか否かマークされる。分割された $B \times B$ ブロックの配列が、次に、例えば、左から右へ、上から下へと、定査され、マークされたブロックが連続して発生するか探索される。マークされたブロックが連続して発生する部分については、微細な走査および形状の適合が遂行される。形状検出器61の構成要素として粗スキャナ65 20を含めることは、オプションであり、システムが用いる計算上の複雑さに依存する。

【0043】微細スキャナ66は、分割され、マークされた、連続する各 $B \times B$ 画素のブロックを、例えば、左から右へ、上から下へと、走査することで、画素の各ライン内の非ゼロの値を持つ第一の画素と、画素の各ライン内の非ゼロの値を持つ最後の画素を検出し、各ラインの検出された非ゼロの最初と最後の画素に、それぞれ、 (x_{start}, y) および (x_{end}, y) なる座標のラベルを付ける。

【0044】形状適合器67は、画素の各ライン上の (x_{start}, y) および (x_{end}, y) なるラベルを付けられた座標を走査する。次に、こうしてラベル付けされた座標領域を、皮膚領域である可能性の高い形状適合器67のメモリ内に格納されたさまざまなサイズおよび縦横比の幾何形状と比較することで、形状の概ねの一致が見られるか決定される。皮膚領域を含む可能性の高い所定の形状、例えば、楕円と、良く適合した形状の輪郭とに一致が見られるか否か決定した後、形状検出器は、良く適合した形状の座標に基づいて、形状検出信号68 40を生成し、これを色調検出器62に供給する。

【0045】いったん形状検出器61によって、皮膚領域である可能性が高いことを示す境界を持つ対象が検出されると、次に、色調検出器62によって、その対象が、皮膚領域に特有な信号エネルギーを持つか否か分析される。ここで使用される信号エネルギーという用語は、ビデオ信号の指定されるグループの画素の輝度（明るさ）の二乗の総和を意味する。信号エネルギーは、直流（DC）信号エネルギーと、交流（AC）信号エネルギーとの二つの成分を含む。対象が、皮膚領域に特有の 50

(9)

特開平10-229505

16

信号エネルギーを含む場合は、色調検出器62は、対象の色調をサンプリングすることで、皮膚色調値のレンジを識別する。色調検出器62は、次に、皮膚色調値の識別されたレンジを、ビデオフレームの残りの部分の色調と比較することで、同一の皮膚色調値を含む他の領域を識別する。

【0046】カラーデジタルビデオ信号は、赤（R）、緑（G）および青（B）のカラー成分を含み、これらは、典型的には、標準のYUVカラービデオフォーマットにて利用可能である。ここで、Yは輝度を表し、UおよびVは両方とも色差（クロミナンス）を表す。輝度（Y）は、ビデオ画像の明るさを表し、クロミナンス（U、V）は、2つのカラー差の値、CrおよびCbを定義する。輝度、二つのカラー差の値CrおよびCb、並びに、3つのカラー成分R、GおよびBの間の関係は、典型的には、以下によって表される：

【数8】

$$Y = 0.299R + 0.587G + 0.114B$$

$$C_r = R - Y$$

$$C_b = B - Y$$

【0047】本発明の一つの実施例においては、色調検出器62は、図7に示すように、皮膚領域検出器70、Crヒストグラム生成器71、Cbヒストグラム生成器72、Crレンジ検出器73、Cbレンジ検出器74、および色調比較器75を含む。

【0048】皮膚領域検出器70は、入力ビデオ信号26と、形状検出器信号68との関係から、形状検出器61によってビデオフレーム内に識別された対象を、 $D \times D$ 画素のブロックに分割する。皮膚検出器60は、好ましくは、識別された形状を、 2×2 画素のブロック（つまり、 $D = 2$ ）に分割し、各画素に対して一つの輝度値を得るため、並びに、各 2×2 画素のブロックに対して一つのCr値と一つのCb値を得る。一例として、図8は、画素100の 4×4 のブロックを示す。輝度値（Y）101は、各画素100に対して存在する。これとは対照的に、 2×2 画素の各ブロックは、一つのCr値102、および一つのCb値103を含む。これは、水平および垂直の両方向の $1/2$ の寸法の位置に存在する。こうして、各 2×2 画素のブロックは、4個の輝度（Y）値101、並びに、一つのCr値102および一つのCb値103を含む。このようなただ一つのCr値とただ一つのCb値を含ませるような分割は、後に説明されるように、識別された対象について皮膚色調サンプリングを遂行する際に重要である。

【0049】皮膚領域検出器70は、識別された対象の周囲に位置する $D \times D$ 画素のブロックが皮膚領域に当たるかを、各 $D \times D$ 画素のブロックが、皮膚領域に特

(10)

特開平10-229505

17

有の信号エネルギーを持つか否かを決定することで分析する。カラービデオ信号の輝度（Y）は、交流（AC）エネルギー成分と、直流（DC）エネルギー成分との二つの信号エネルギー成分を持つ。皮膚領域の画素は、典型的には、指定された閾値エネルギー T_{en} より小さな値のACエネルギー成分を持つ。

【0050】皮膚領域は、カラービデオ信号の輝度（Y）に対するACエネルギー成分の計算に基づいて検出される。ITU-T勧告H. 263（「狭帯域チャネル」

$$F(u,v) = C(u)C(v) \sum_{i=0}^{u-1} \sum_{j=0}^{v-1} f(i,j) \cos \frac{(2i+1)u}{4} \pi \cos \frac{(2j+1)v}{4} \pi \quad (1)$$

ここで、 $F(u, v)$ は、離散コサイン変換（DCT）関数を表し、 $C(u)$ および $C(v)$ は、以下のように定義される：

【数10】

$$C(\omega) = 1/\sqrt{2} \quad \text{対して } \omega = 0$$

$$C(\omega) = 1 \quad \text{対して } \omega = 1, 2, 3, \dots$$

【0051】これらが、 $D \times D$ 画素のブロックの各画素位置 $F(u, v)$ に対して総和される。次に、AC信号エネルギー $E(m, 1)$ が、式（2）に示すように、離散コサイン変換関数 $F(u, v)$ の二乗から、直流電流（DC）信号エネルギー $F(m, 1(0, 0))$ の二乗を引くことによって決定される：

【数11】

$$E_{m,j} = \sum_{u=0}^{M-1} \sum_{v=0}^{M-1} F_{m,j}(u,v)^2 - F_{m,j}(0,0)^2 \quad (2)$$

【0052】次に、AC信号エネルギー $E(m, 1)$ が、閾値エネルギー T_{en} と比較される。以下に示すように、 $D \times D$ 画素の各ブロックに対して、AC信号エネルギー $E(m, 1)$ が、前もって選択された閾値エネルギー T_{en} より小さな場合は、画素のそのブロックは、皮膚領域であるものと識別される：

【数12】

$$E(m,i) < T_{en} \quad \text{皮膚領域}$$

$$E(m,i) \geq T_{en} \quad \text{非皮膚領域}$$

【0053】典型的には、 $D \times D$ 画素のあるブロックが120、000以下のAC信号エネルギー値を持つ場合は、その画素のブロックは、皮膚領域であるものと識別される。輝度パラメータの信号エネルギー成分を用いることは、皮膚でない領域は皮膚領域よりも高い信号エネルギー成分を持つ傾向があるので皮膚領域を決定するのに有効である。皮膚でない領域を識別し、これら領域をカラーサンプリングプロセスから除外することで、サンプリングされた画素のカラーが、実際に皮膚領域の画素である確率が増し、サンプリングされるべき色調レンジ

18

＊ルに対するビデオ符号化）において説明される離散コサイン変換（DCT）技法を含む方法が、輝度（Y）の信号エネルギーを計算するために使用することができる。一つの実施例においては、 $D \times D$ 画素の各ブロックに対する輝度のACエネルギー成分およびDCエネルギー成分の決定に当たっては、最初に、各画素に対する式（1）に示される離散コサイン変換（DCT）関数、 $F(u, v)$ が、計算される：

【数9】

の精度が向上する。

【0054】いったん皮膚領域検出器70によって、ある $D \times D$ 画素のブロックが、皮膚領域として識別されたら、その $D \times D$ 画素のブロックのCr値およびCb値が、それぞれ、Crヒストグラム生成器71およびCbヒストグラム生成器72によってサンプリングされる。前に述べたように、 $D \times D$ 画素のブロックを、 2×2 画素のブロックとすることは、これらのブロックが、ただ一つのCr値と、ただ一つのCb値のみを含むために利点がある。次に、Crヒストグラム生成器71およびCbヒストグラム生成器72は、それぞれ、サンプリングされたCr値およびCb値のヒストグラムを生成する。

【0055】CrヒストグラムおよびCbヒストグラムの生成を終えたら、サンプリングされた対象の皮膚色調を表すカラー色調のレンジが、Crレンジ検出器73およびCbレンジ検出器74によって、統計的分析技法を使用して決定される。例えば、CrおよびCbの平均値およびモード値が、各データセットと共に、サンプリングされた $D \times D$ 画素のブロックに対して決定される。これらCrおよびCbの平均値およびモード値が、互いに、ある指定される距離 D_p 以内にある場合は、それらCrおよびCbの平均値およびモード値は、単一のピークを表すものと識別される。その後、 $D \times D$ 画素の各ブロックについて、画素のカラー色調が、所定の距離、例えば、単一のピークを表すCrおよびCbの平均値およびモード値からの1標準偏差の距離内にあるか否かが決定され、ある場合は、その画素のカラー色調が、皮膚色調のレンジ内に含まれる。他方、平均値およびモード値が、指定される距離 D_p より大きな距離にある場合は、CrおよびCbのそれぞれの平均値およびモード値は、2つの別個のピークを表すものと識別される。そして、2つの別個のピークを表すCrおよびCbの平均値およびモード値を持つ $D \times D$ 画素のブロックの画素のカラー色調は、皮膚色調のレンジ内に含まれない。

【0056】Crレンジ検出器73およびCbレンジ検出器74によって生成されたCr値およびCb値のレンジに基づいて、色調比較器75は、入力ビデオ信号26の全フレームを分析することで、同一のクロミナンス値を持

19

つ他の全ての画素を検出する。他の画素が検出されると、色調比較器75によって、そのような皮膚領域の検出を示す皮膚情報信号76が生成され、フィルタセレクタ90に供給される。

【0057】皮膚検出器60は、上に説明の分析を、ビデオシーケンスの各フレームに対して遂行することも、あるいは、オプションとして、単一のフレームを分析することもできる。いずれにしても、次に、色調比較器75は、皮膚色調値のこうして得られたレンジを用いて、後続の所定の数のフレーム内の皮膚領域を識別する。

【0058】オプションとして識別される画像パラメータが目鼻口（ENM）領域である本発明の一つの実施例においては、画像パラメータ抽出器50は、図3に示すように、目鼻口（ENM）領域検出器80を含む。ENM領域検出器80は、最初に、ビデオシーケンス内の対象の形状を分析することにより、顔である可能性の高い対象を検出する。顔である可能性が高い対象をさらに分析することで、その対象が、目鼻口（ENM）領域を含むか否か識別される。ENM領域検出器80は、図9に示すように、形状検出器81、および目鼻口（ENM）

検出器82を含む。

【0059】ENM領域検出器80の形状検出器81は、ビデオフレーム内の顔を含む可能性が高い対象を、対象のエッジの識別、並びに、そのようなエッジが所定の形状の輪郭に類似するか否かに基づいて検出する。例えば、楕円は、人の顔あるいは頭の形に類似するため、ビデオシーケンスを分析することで、対象の楕円に類似する輪郭が識別され、これによって、ビデオシーケンス内の目鼻口領域を含む可能性が高い幾つかの領域が検出される。

【0060】形状検出器81は、図6に示す形状検出器61と同様の幾つかの構成要素を含む。形状検出器81の機能およびこれを含む構成要素は、前述の形状検出器61と同一であり、ENM領域の検出を行なうために必要とされる全ての形状検出情報は、形状検出器61との関連で上に説明されたのと同一の方法で決定されるものとする。さらに、形状検出器81を、ENM領域検出器80の構成要素として含めるか否かは、画像パラメータ抽出器50が、皮膚検出器、例えば皮膚検出器60、およびENM領域検出器、例えばENM領域検出器80の両方を含む場合は、形状検出情報を皮膚検出器60の形状検出器61から得ることも可能であるために、オプションとされる。

【0061】図9に戻り、形状検出器81によって識別された（あるいはオプションとして皮膚検出器60の形状検出器61によって識別された）一つあるいは複数の対象の輪郭が楕円と良く適合する場合は、この形状検出情報が、目鼻口（ENM）領域検出器82に供給される。次に、ENM領域検出器82は、対称ベースの技法を用いて、その良く適合した楕円が、目鼻口領域を持つ

(11)

特開平10-229505

20

か否か決定する。ENM領域検出器82は、探索領域識別器83および探索領域スキャナ84を含む。

【0062】ENM領域検出器82は、良く適合した楕円の座標を受信し、識別された対象を、長方形のウィンドウと、楕円領域（長方形ウィンドウ内に位置しない識別された対象の残りの部分を含む領域）とに分割する。一例として、図10は、長方形ウィンドウ85と、楕円領域86（長方形ウィンドウ85内に位置しない楕円の残りの部分を含む領域）とに分割された良く適合した楕円の輪郭を示す。ENM領域検出器82は、対称ベースの分析を用いて、長方形ウィンドウ、例えば、長方形ウィンドウ85を検出し、目、鼻、および口領域に対応する楕円の領域を捕捉する。

【0063】ENM領域検出器82は、長方形ウィンドウ、例えば長方形ウィンドウ85を検出するための探索領域を、探索領域識別器83を用いて決定する。これには、図10に示すような楕円輪郭の中心点（ x_0 , y_0 ）の座標が、長方形ウィンドウ、例えば長方形ウィンドウ85の位置を知るために用いられる。

【0064】ENM領域の中心点を見つけるための探索領域は、典型的には、 $S \times T$ 画素のサイズ、例えば 12×15 画素のサイズの長方形とされ、好ましくは、楕円輪郭の主軸および副軸に対して固定されたサイズを持つように選択される。この明細書において使用される主軸という用語は、図10の楕円輪郭との関連で定義され、点 y_1 と y_2 の間の楕円を二分する線分を意味する。この明細書において使用される副軸も図10の楕円輪郭との関連で定義され、これは、点 x_1 と x_2 の間の楕円を二分する線分を意味する。一例として、楕円は、主軸に沿って50画素の長さを持ち、副軸に沿って、30画素の長さを持つものと想定する。長方形ウィンドウのサイズは、好ましくは、 25×15 画素のサイズを持つように選択される。このサイズは、主軸および副軸の両方に沿っての楕円の長さおよび幅の概ね二分の一に等しく、形状を検出するための目鼻口（ENM）領域の最もありそうな位置を捕捉する。

【0065】楕円内の長方形ウィンドウ、例えば、長方形ウィンドウ85が楕円の内にあることを検出したら、探索領域スキャナ84によって、長方形ウィンドウが分析され、各候補位置が、対象の目鼻口領域に特有な、対称の軸を持つか調べられる。対称の軸は、対象を二つの等しい部分に分割する線分である。対称的に描画される顔および頭は、目の間、鼻の中央、口の半分を通る、垂直な対称の軸を持つ。各半分は、おのおのが、一つの目、鼻の半分、および口の半分を含むために対称である。例えば、図10に示すように、探索領域スキャナ84は、左から右へと、長方形ウィンドウ85内の画素の垂直な各列を、主軸に平行な線分を使用して選択し、目の間、鼻の中央、口の半分を通して位置する対称の軸を探索する。

50

(12)

特開平10-229505

21

【0066】顔の対称の軸を決定した後、ENM領域検出器82は、長方形ウィンドウ、例えば、長方形ウィンドウ85の目鼻口領域と関連する画素の座標に対応するENM領域信号87を生成する。ENM信号87は、フィルタセクタ90に対して、楕円の目鼻口領域と関連する画素を知らせる。

【0067】目鼻口領域の検出は、例えば、ビデオ会議状況においてしばしば見られるように、対象がカメラに真っすぐに向っていない場合に、影響を受けることがある。ENM領域検出器80は、対象が真っすぐにカメラに向っていない場合や、対象の顔に髪がかかっていたり、対象が眼鏡を掛けている場合でも、入力ビデオ画像の目鼻口領域の検出を可能にする機構を持つ。ENM領域検出器80は、このために、鼻を通り、口を横断する縦軸についての顔の形状の典型的な対称性を利用する。対称の軸は、図10に示す楕円上に示される垂直軸yに対して角度 θ だけ傾斜することがある。このような傾斜した楕円に対しては、長方形ウィンドウ85が、ウィンドウの中心のまわりを傾斜角度値だけ回転され、こうして、傾斜している場合でも、目鼻口領域の検出ができるようにされる。好ましくは、角 θ は、 -10 から $+10$ 度の範囲内の値を持つようにされる。

【0068】オプションとして識別される画像パラメータが移動する対象の位置の追定であるような本発明の一つの実施例においては、画像パラメータ抽出器50は、図3に示すような移動領域検出器105を含む。移動領域検出器105は好ましくは、移動する対象と関連する画素を、ブロックベースの移動推定技法を使用して識別する。ブロックベースの移動推定技法においては、ビデオシーケンスの複数のフレームが画像ブロックに分割され、次に、ビデオシーケンスの前のフレーム内の画像ブロックを分析することで、現在のフレーム内のこれら同一の画像ブロックの位置が予測される。前の画像ブロック内に識別された画素強度の位置から現在の画像ブロック内のこれら同じ識別された画素強度のシフトされた位置までの間の変位の大きさによって、ビデオシーケンスのフレーム内のある対象の移動領域が決定される。移動領域検出器105は、図11に示すように、一時ダウンサンブラ106および移動検出器107を含む。

【0069】一時フィルタ106は、移動検出の対象となるビデオ信号のフレームの数を、入力ビデオ信号26として供給されるフレームの数を制限することによって、制限する。一時ダウンサンブラ106は、図4に示す前置処理回路52の一時ダウンサンブラ53と同様の構成要素を含む。一時ダウンサンブラ106の機能は、前に説明した一時ダウンサンブラ53と同一であり、移動検出を遂行するためのフレームの数の制限は、前に一時ダウンサンブラ53との関連で説明されたのと同じの方法で行なわれるものとする。さらに、移動領域検出器105の構成要素として一時ダウンサンブラ106を含

22

めるか否かは、画像パラメータ抽出器50が、エッジ検出器、例えば、皮膚検出器60を含む場合は、フレームの数の制限は、エッジ検出器51の前置処理回路52の一時ダウンサンブラ、あるいは、皮膚検出器60の形状検出前置プロセッサ64の一時ダウンサンブラによっても達成できるために、オプションである。

【0070】一時ダウンサンブラ106から得られた（あるいはオプションとしてエッジ検出器51の前置処理回路52の一時ダウンサンブラ、あるいは皮膚検出器60の形状検出前置プロセッサ64から得られた）現在のビデオフレーム内の制限された数のフレームは、移動検出器107に供給される。移動検出器107は、現在のフレームに対する入力信号を、中心点 (m_0, r_0) を持つ、 $M \times R$ 画素のサイズ、例えば、 16×16 画素のサイズの画像ブロックに分割する。その後、移動検出器107は、ビデオシーケンスの前のフレームに対する $M \times R$ の画像ブロックの位置に対応する中央点 (m_0, r_0) を検出し、 $(M \pm p) \times (R \pm p)$ なるサイズの画像ブロックに対応する画素の領域を分析する（ここで、 p は、現在のフレーム内に表された $M \times R$ なる画像ブロックに対して増加される画素の数を表す）。この領域は、分析されるべき前のフレーム内の画像ブロックのサイズを定義する。一例として、図12には、ビデオシーケンスの現在のフレーム111に対する $M \times R$ 画素なるサイズのブロック110と、ビデオシーケンスの前のフレーム116に対する $(M \pm p) \times (R \pm p)$ 画素なるサイズの対応するブロック120が示される。

【0071】移動検出器107は、現在のフレーム111に対する $M \times R$ なるサイズの画像ブロック110の画素強度を走査することで、対象の画素の一致を推定する。移動検出器107は、さらに、前のフレーム116に対する $(M \pm p) \times (R \pm p)$ なる画像ブロック120の画素強度を走査することで、対象の画素の一致を推定する。移動検出器107は、ビデオシーケンスの両方のフレーム内の画素強度を、例えば、左から右へ、あるいは上から下へ、走査し、次に、これらを互いに比較することで、任意の対象の一致する画素の変位の大きさを測定する。

【0072】変位の大きさは、ビデオシーケンスの現在のフレームにおける移動領域の画素位置を定義する。例えば、図12には、現在のフレーム111の $M \times R$ なる画像ブロック110内に、円112の輪郭が示される。同様に、前のフレーム116の画像ブロック120内には、円112と同一サイズの円114の輪郭が示される（中央点 (m_0, r_0) に対応する円112の位置についても点線にて示される）。円114の位置と、円112の位置の比較から、円114に対する中央点 (m_1, r_1) の位置が、円112の中央点 (m_0, r_0) から距離 d だけシフトしたことがわかる。この距離 d は、二つの円112と114の間の変位の大きさを表し、対象の移

(13)

特開平10-229505

23

動領域に対する画素位置を定義する。このような移動領域が検出された場合、このような移動領域に対する画素位置を示す移動領域情報信号117が移動領域検出器105によって生成され、フィルタセクタ90に供給される。

【0073】また、ビデオ画像の背景領域についても画像パラメータ抽出器50によって識別することが可能である。背景領域は、画像パラメータ抽出器50によって識別される最後の画像パラメータであり、前に識別された任意の画像パラメータ領域、例えば、皮膚領域、目鼻口領域、および移動領域に含まれないビデオ画像の画素位置を表す。例えば、ビデオシーケンスが、ビデオカメラに
10 対面している人を含むものと想定する。画像パラメータ抽出器、例えば、画像パラメータ抽出器50が、対象の皮膚領域およびエッジのみを識別するようにプログラムされている場合は、皮膚画素として識別されず、かつ、ビデオカメラに
20 対面する人の輪郭（エッジ）をも含まない全ての画素は、画像パラメータ抽出器50によって、背景領域の画素を表すものと決定される。そして、背景領域の位置が、画像パラメータ抽出器50によって、画像パラメータ検出信号100の一部分として、フィルタセクタ90に供給される。

【0074】図2に再び戻り、いったん画像パラメータ抽出器50が上で説明された要領で画像パラメータと関連する画素を識別すると、フィルタセクタ90は、各識別された画像パラメータのビデオシーケンスの内容全体からみた重要度をランク付けする。その後、各識別された画像パラメータのランクに基づいて、フィルタセクタ90は、フィルタ95の強度を調節するためのフィルタリング係数を決定する。その後、フィルタ95は、
30 各識別された画像パラメータと関連する画素をフィルタリングするために用いられる。

【0075】フィルタセクタ90は、好ましくは、各識別された画像パラメータのビデオシーケンスの内容からみた重要度を、知覚的意味論に基づいてランク付けするようにプログラムされる。ここで使用される知覚的意味論という用語は、ビデオ画像内の主題の視聴者の焦点に従う差別化を意味する。こうして、フィルタセクタ90は、画像パラメータ抽出器50によって識別されたビデオシーケンス内の画像パラメータを、視聴者の最も
40 あり得る焦点に基づいて、ランク付けする。例えば、ビデオ会議の状況においては、ビデオシーケンスは、典型的には、部屋の中にカメラに
50 対面する一人あるいは複数の人物を含む。このような画像の視聴者は、彼あるいは彼女の注意を、部屋の中の一人あるいは複数の人物、特に、人物の頭部を取り巻く背景領域ではなく、彼らの顔および両手に向ける傾向があるために、顔の目鼻口（ENM）領域と関連する画素には、皮膚領域と関連する画素より重要なランクを与えられる。同様に、顔の皮膚領域と関連する画素には、人の頭部を取り巻く背景と関連

24

する画素よりも高いランクを与えられる。

【0076】図13は、フィルタセクタ90によって実行される識別された画像パラメータのグループの一例としてのランク付けを示す。図13に描画されているビデオ画像は、部屋の中のカメラに向っている人物を示す。ここでは、画像パラメータ抽出器50によって識別された画像パラメータは、目鼻口（ENM）領域、皮膚領域、エッジ領域、移動領域、および背景領域を含む。フィルタセクタ90は、この例に示されるような識別された画像パラメータを、このビデオシーケンスを視聴する視聴者が向けると考えられる焦点の順に従ってランク付けする。このような画像を視聴する者は、彼あるいは彼女の注意を部屋の中の人物、特に、その人物の顔に向ける傾向があるために、目鼻口（ENM）領域の内側に位置するエッジと関連する画素には、他の任意の画像パラメータと関連する画素より重要なランクが与えられる。同様に、皮膚領域と関連する画素には、移動領域と関連する画素より重要なランクが与えられ、背景領域と関連する画素には、最低のランクが与えられる。

【0077】画像パラメータ抽出器50によって識別された画像パラメータを、それらの相対的な重要度に従ってランク付けした後に、フィルタセクタ90は、各識別された画像パラメータと関連する画素をフィルタリングするフィルタ95の強度を調節するためのフィルタリング係数Nを決定する。フィルタの強度は、フィルタ、
例えば、フィルタ95によってフィルタされる周波数のレンジを定義する。弱いフィルタ強度は、フィルタに対して、広いレンジの周波数を定義し、このために、ビデオシーケンスからは、非常に少量の画像情報（バイト）が除去される。反対に、強いフィルタ強度は、フィルタに対して、狭いレンジの周波数を定義し、この結果、ビデオシーケンスからは多量の画像情報（バイト）が除去される。

【0078】フィルタセクタ90は、好ましくは、フィルタ係数Nの選択において、ビデオシーケンスの内容にとって最も重要であるとランク付けされた画像パラメータと関連する画素には、最も弱いフィルタ強度が適用されるように選択し、こうして、重要な画像情報が除去されないようにする。例えば、図13において、ビデオシーケンスの内容にとって最も重要であるとランク付けされたENM領域の内側のエッジと関連する画素には、これらの画素を最も弱いフィルタ強度にてフィルタリングするように指定するフィルタ係数が割当てられる。これは、弱いフィルタ強度は、フィルタに対して、広いレンジの周波数を定義するために、これらの画素からは、非常に少量の画像情報（バイト）のみが除去されることを意味する。反対に、ビデオシーケンスの内容にとって最も重要でないとしてランク付けされた背景領域と関連する画素には、これらの画素を最も強いフィルタにてフィルタリングするように指定するフィルタ係数が割当てられる。

(14)

特開平10-229505

25

26

【0079】フィルタ係数Nは、典型的には、 $1 < N < 256$ として定義される数値を持つ。Nの値が大きくなるとは、より弱いフィルタを意味し、一方、Nの値が小さくなるとは、より強いフィルタを意味する。Nが256より大きな場合は、フィルタ係数（つまり、後に説明の $1/(2+N)^2$ ）が、非常に大きな数となり、小さな数を計算するための精度を超えるために、フィルタ係数に*

テーブル1

領域	C I F *	Q C I F **
エッジENN領域	16	64
非エッジENN領域	8	32
エッジ皮膚領域	8	32
非エッジ皮膚領域	2	8
エッジ移動領域	4	16
非エッジ移動領域	2	8

* C I Fは、ビデオ信号を送送するための標準画像フォーマットで

ある共通中間フォーマット (common intermediate format) を表す。

** Q C I Fは、四分の一 (quarter) C I F (共通中間フォーマット)

を表す。

【0080】いったんフィルタ係数Nがフィルタセクタ90によって決定されたら、識別された画像パラメータと関連する画素が、フィルタ95を用いてフィルタされる。フィルタ95は、好ましくは、以下の形式のフィルタとされる：

【数14】

$$F = \frac{1}{(2+N)^2} \begin{bmatrix} 1 & N & 1 \\ N & N^2 & N \\ 1 & N & 1 \end{bmatrix}$$

$$= \frac{1}{(2+N)^2} \begin{bmatrix} 1 & N & 1 \\ 1 & N & 1 \end{bmatrix} \times \begin{bmatrix} 1 & N & 1 \\ 1 & N & 1 \end{bmatrix}$$

ここで、Nは、フィルタリング係数である。その後、各識別された画像パラメータと関連する画素が、フィルタセクタ90によって決定されたフィルタリング係数を用いてフィルタ95によってフィルタされ、フィルタされたビデオ信号は、ビデオコーデック10のソース符号器32に符号化のために供給される。

【0081】対象指向のフィルタ12は、上に説明された分析を、ビデオシーケンスの各フレームに対して実行することも、あるいはオプションとして単一のフレームを分析することもできる。いずれにしても、次に、フィルタ95が、これらフレームに対して選択されたフィルタリング係数を用いて、ビデオシーケンスの指定された

* 対する256を超える値は適当でない。例えば、 $N=256$ の場合は、係数 $1/(2+N)^2$ は、 $1/(2+256)^2 = 1.5 \times 10^{-5}$ となるが、これは、非常に大きな数である。選択された画像パラメータに割り当てられるフィルタリング係数の一例をテーブル1に示す。

【数13】

数の後続のフレームをフィルタリングする。

【0082】オプションとして、対象指向のフィルタ12は、ビデオ符号器/復号器（コーデック）、例えば、ビデオコーデック10との関連で使用することも可能である。以下においては、図1に示すビデオコーデック10の他の構成要素との関連で、対象指向のフィルタ12の動作について説明する。ビデオコーデック10は、ビデオ符号器22およびビデオ復号器24を含む。ここで、ビデオ符号器22は、符号化コントローラ16によって制御される。

【0083】符号化動作に対しては、ビデオコーデック10は、入力ビデオ信号26を受信し、これを対象指向のフィルタ12に送る。対象指向のフィルタ12は、入力ビデオ信号26を上記説明したようにフィルタし、次に、フィルタされたビデオ信号をソース符号器32に送る。ビデオ符号器22は、フィルタされたビデオ信号を、符号化コントローラ16の制御下で符号化し、符号化された出力ビット流30を生成する。一つの実施として、ビデオ符号器22は、フィルタされたビデオ信号を、ソース符号器32、ビデオ多重符号器34、送信バッファ36、および送信符号器38を用いてエンコードし、符号化された出力ビット流30を生成する。

【0084】復号動作においては、ビデオコーデック10は、符号化された入力ビット流40を受信する。ビデオ復号器24は、符号化された入力ビット流40を、受信復号器42、受信機バッファ44、ビデオ多重復号

(15)

特開平10-229505

27

28

器46、およびソース復号器48を使用してデコードし、ビデオ出力信号50を生成する。本発明が一つの実施例との関連で説明されたが、当業者においては、勿論、他の構成も明白であることを理解できるものである。

【図面の簡単な説明】

【図1】本発明の原理の一例との用途を具現するビデオ符号器／復号器（コーデック）のブロック図を示す図である。

【図2】本発明による対象指向のフィルタのブロック図を示す図である。

【図3】図2の画像パラメータ抽出器のブロック図を示す図である。

【図4】図3のエッジ検出器のブロック図を示す図である。

【図5】図3の皮膚検出器のブロック図を示す図である。

【図6】図5の形状検出器のブロック図を示す図である。

【図7】図5の色調検出器のブロック図を示す図である。

【図8】4×4画素のブロック図を示す図である。

【図9】図3の目鼻口（ENM）領域のブロック図を示す図である。

*【図10】指円内に位置する長方形のウィンドウを示す図である。

【図11】図3の移動領域検出器のブロック図を示す図である。

【図12】移動領域検出器のブロックベース移動補償（給定）技法を示す図である。

【図13】ビデオシーケンス全体に対する画像パラメータの重要度に基づく画像パラメータのランク付けの一例を示す図である。

【符号の説明】

10 ビデオコーデック（符号器／復号器）

12 対象指向のフィルタ

16 符号化コントローラ

22 ビデオ符号器

24 ビデオ復号器

26 入力ビデオ信号

32 ソース符号器

34 出力ビット流

40 符号化された入力ビット流

42 受信機復号器

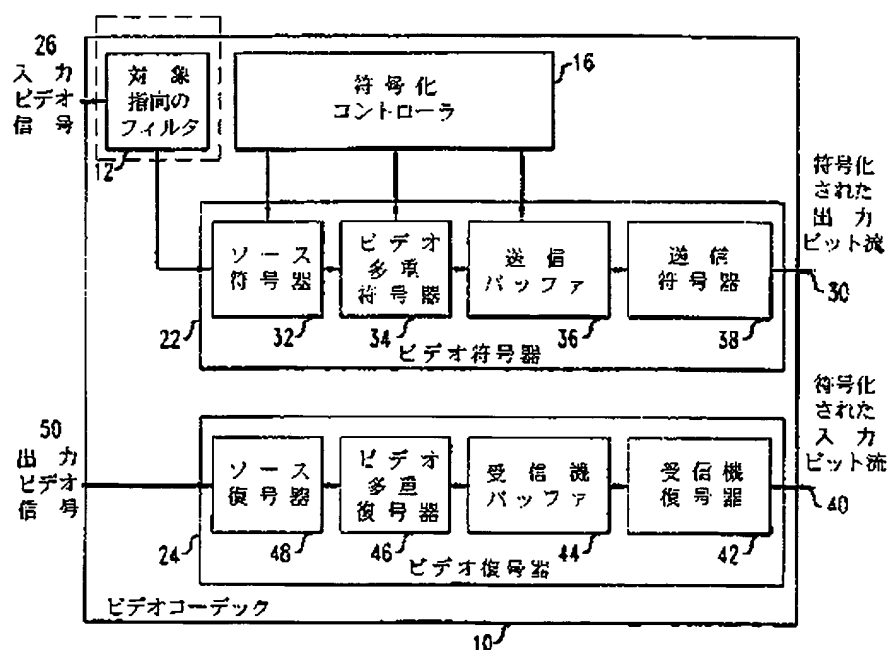
44 受信機バッファ

46 ビデオ多重復号器

48 ソース復号器

* 50 ビデオ出力信号

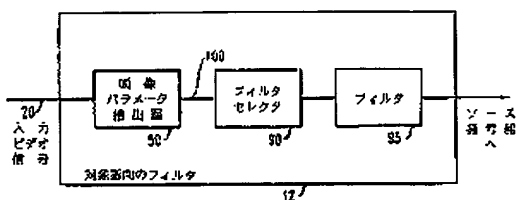
【図1】



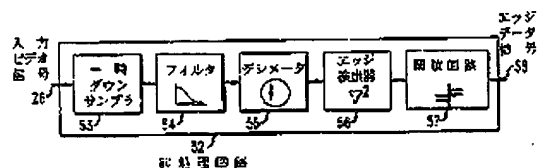
(16)

特開平10-229605

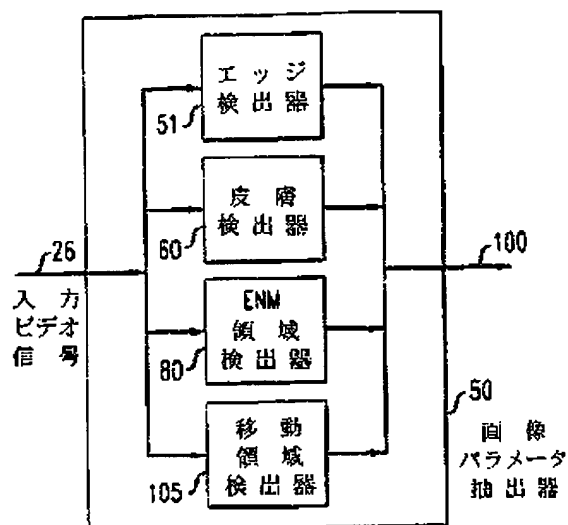
【図2】



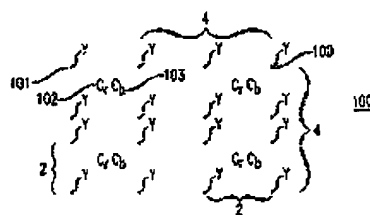
【図4】



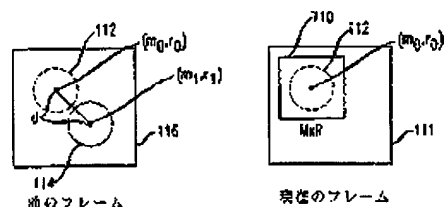
【図3】



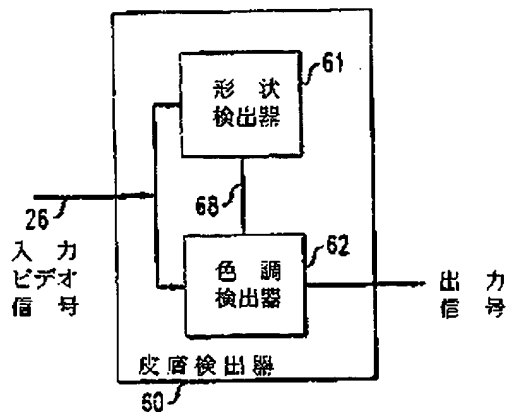
【図8】



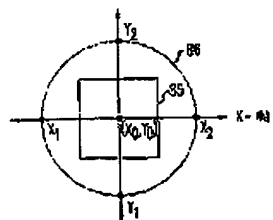
【図12】



【図5】



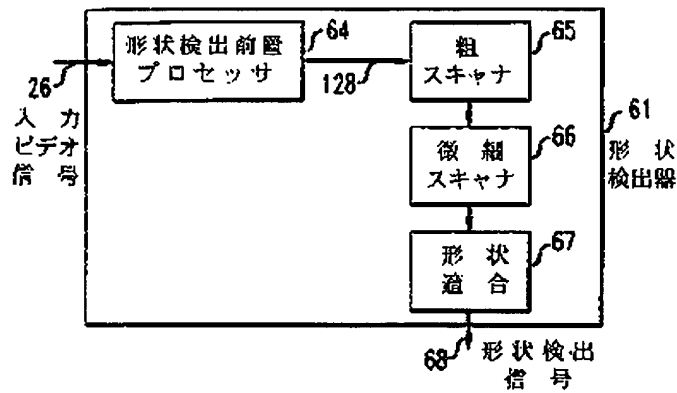
【図10】



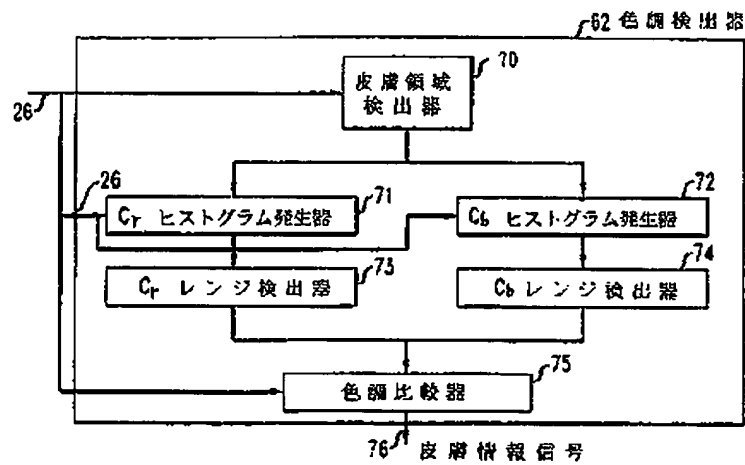
(17)

特開平10-229505

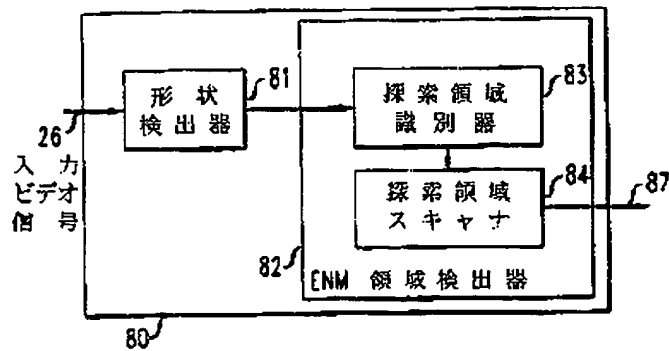
【図6】



【図7】



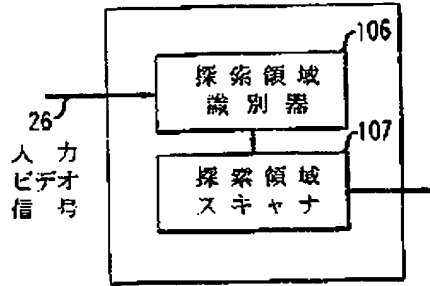
【図9】



(18)

特開平10-229505

【図11】



【図13】

